Exploration of the Shared Gene Signatures and Molecular Mechanisms Between Breast Cancer and Endometriosis

Linyue Hai
1 Tianjin Medical University Cancer Institute & Hospital, Tianjin Medical University

Xuchen Cao
1 Tianjin Medical University Cancer Institute & Hospital, Tianjin Medical University

Chunhua Xiao (✉ yuepu1429@163.com)
1 Tianjin Medical University Cancer Institute & Hospital, Tianjin Medical University

Research Article

Keywords: breast cancer, endometriosis, bioinformatics, genetic signature, immune cell infiltration

Posted Date: March 27th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2649506/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background

Globally, breast cancer (BC) is the most common type of cancer and the second leading cause of cancer-related deaths among women. Early diagnosis and survival outcomes will be improved through the identification of modifiable risk factors and the development of better disease management strategies. There are many similarities between endometriosis and breast cancer, both in terms of risk factors and developmental characteristics. Thus, it would be beneficial to explore the common mechanisms behind the onset of BC and endometriosis to develop more effective intervention strategies in the future. In this study, bioinformatics was used to explore the key molecules and pathways that mediate the co-occurrence of BC and endometriosis.

Method

From the Gene Expression Omnibus (GEO) database, datasets for BC (GSE15852) and endometriosis (GSE5108) were downloaded. By using the GEO2R online tool, we were able to identify the differentially expressed genes (DEGs) between two diseases. Afterward, a protein-protein interaction network (PPI) was constructed based on DEG enrichment analysis. Additionally, the hub genes were identified using the STRING database and Cytoscape software. We investigated the relationship between hub gene expression levels and clinical expression, pathological stage, age, and prognosis. As a final step, transcription factor interaction, stemness score, and immune cell infiltration analysis were conducted on hub genes in BC.

Results

We identified 33 overlapping DEGs (18 downregulated genes and 15 upregulated genes) for further analysis. The significant functional pathways of DEGs were enriched in regulating the pluripotency of stem cells and the mis-regulation of transcription in cancer. Additionally, five key hub genes were identified, including HOXA10, PAX8, MSX1, FGFR1, and INHBA. Pathological stages, age, stemness score, and immune infiltration were associated with the expression level of hub genes.

Conclusion

A novel insight into the molecular mechanism of endometriosis complicated with BC is provided by the finding that HOXA10, PAX8, MSX1, FGFR1, and INHBA were hub genes for the co-occurrence of BC and endometriosis.

Introduction
It is estimated that 281,550 new cases and more than 43,600 deaths will occur in 2021 due to breast cancer (BC), the most common type of cancer among women worldwide. Global cancer statistics indicate that BC accounts for 30% of female cancer deaths.\cite{1,2} When breast cancer spreads to bones, liver, lung, and brain, it is considered the most severe form of breast cancer.\cite{3} Advanced metastatic breast cancer significantly reduces the likelihood of survival. Therefore, the identification of modifiable risk factors and the development of better disease management strategies will contribute to early diagnosis and improved survival rates. Breast cancer incidence varies significantly from population to population. According to existing research, age, obesity\cite{4}, alcohol abuse\cite{5}, family history, benign breast disease \cite{6}, and mutations of the BRCA gene all contribute to the development of breast cancer.\cite{7} It is imperative to identify those individuals at a high risk of developing breast cancer.

The disease process of endometriosis is characterized by hormonal and inflammatory changes.\cite{8,9} Dysmenorrhea, non-periodic pelvic pain, dysuria, difficulty defecating, and other symptoms are caused by endometriosis, seriously affecting the quality of life of patients.\cite{10–12} The growth pattern of endometriosis is similar to that of cancer, with distant metastases, a systemic inflammatory milieu, and resistance to apoptosis.\cite{13–15} Endometriosis is associated with an increased risk of malignant tumors, such as ovarian cancer, endometrial cancer, and breast cancer.\cite{16} It appears that endogenous estrogen, reproductive characteristics, obesity, and the use of hormone replacement therapy contribute to the development of endometriosis, as well as BC. Researchers are interested in researching the relationship between endometriosis and breast cancer both in terms of risk factors and developmental characteristics. There has also been evidence that endometriosis increases BC risk.\cite{17–20} Although there is clinical and epidemiological evidence that endometriosis increases BC risk, little is understood about their shared gene expression signatures. Therefore, it may be beneficial to explore the common initiation mechanism of BC and endometriosis in order to develop strategies and interventions for future treatment.

Several diseases have been studied using microarrays and bioinformatics methods to identify new biomarkers to improve diagnosis and treatment.\cite{21,22} This study aims to identify genes that are associated with endometriosis and BC. It is the first study that has explored shared genetic signatures between BC and endometriosis using a systems biology approach and is expected to provide new insights into the biological mechanisms of both diseases.

**Materials And Method**

**Gene Expression Profile Data Collection**

For the purpose of identifying the genes associated with BC and endometriosis, two microarray datasets (GSE5108 and GSE15852) were screened from the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo ).\cite{23,24} Screening criteria included: (1) Homo sapiens Expression Profiling by array; (2) All tissues were pathologically supported; (3) the gene expression profiling must include both cases and the paired controls; (2) these datasets must provide processed data or raw data
that can be used for re-analysis; (5) the patients were not treated with drugs or hormones. GSE5108 included 11 patients with ectopic and eutopic endometrium, while GSE15852 included 43 patients with breast cancer and their matched normal controls.

**Differentially Expressed Genes Selection**

To obtain differentially expressed genes (DEGs), the original file downloaded from the GEO database is processed using the GEO2R package (https://www.ncbi.nlm.nih.gov/geo/geo2r/) based on R software. Using GEO2R, data can be read and multiple differential expressions calculated based on two R packages (GEOquery and Limma). By comparing gene expression profiles between the two conditions, DEGs can be identified. The folding changes (FCs) of individual genes were calculated. \(|\log_2 FC| \geq 1 \) and \( p < 0.05 \) was the DEGs specific cutoff criteria for DEGs. Additionally, Venn diagrams were used to visualize the DEGs that overlap between GSE5108 and GSE15852.

**Functional Annotation and Enrichment Analysis**

A wide range of biological functions and significance are evaluated using enrichment analysis of genes. Gene ontology (GO) categories and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were conducted in the DAVID Bioinformatics Resources (http://david.abcc.ncifcrf.gov/) using \( p < 0.05 \) as a cutoff value. The GO functional analysis of the integrated differential genes was divided into the following three sections: biological process (BP), molecular function (MF), and cell component (CC). Moreover, Metascape (http://metascape.org/) is a powerful gene function annotation tool that enables users to apply popular bioinformatic analysis methods to batch gene and protein analysis to gain an understanding of gene and protein function. All genes in the given gene list were used for enrichment purposes. The items with a \( p \)-value of 0.01, a minimum count of three, and an enrichment factor greater than 1.5 were collected and grouped according to their similarity as members.

**Gene Expression and Survival Prognostic Analysis**

The DEG differential expression level was determined using UALCAN (http://ualcan.path.uab.edu/analysis.html), a website that provides a variety of network resources. There is an online website, Kaplan Meier (KM) Plotter (http://kmplot.com/analysis/), which combines GEO, TCGA, and EGA datasets with 54000 genes in 24 tumors, allowing one to assess the impact on the survival rate. By using KM Plotter, we were able to obtain the expression and survival prognosis data of DEGs in BC.

**Protein-protein Interaction (PPI) Network Analysis**

To further investigate the interactions between these common genes, we constructed the PPI network using the Interaction Gene Retrieval Database Search Tool (STRING) (https://cn.string-db.org/) and conducted visual processing analysis using Cytoscape 3.9.1. To identify key functional genes and obtain a cluster score, the minimum Common Oncology data element (MCODE) in Cytoscape tools was used (selection criteria: degree critical = 2; node score critical = 0.2; k core = 2; maximum depth = 100).
Following this, we used the CytoHubba tools of Cytoscape to identify important genes in PPI by identifying the network characteristics of nodes and measuring nodes. As part of the scoring process, we used MCC (Maximum Clique Centrality), DMNC (Density of Maximum Neighborhood Component), MNC (Maximum Neighborhood Component), Degree, EPC (Edge Percolated Component), BottleNeck, EcCentricity, Closeness, Radiality, and Betweenness to score all node genes, and ten of the top nodes were selected as the hub genes. GeneMANIA (http://www.genemania.org) is a flexible, user-friendly web interface for generating hypotheses about gene function, analyzing gene lists, and prioritizing genes for functional assays. A co-expression network of hub genes was then constructed using GeneMANIA (http://www.genemania.org/).[34]

**Immunohistochemistry (IHC) Staining**

The human protein mapping (https://www.proteinatlas.org/) provides information on the expression and location of various proteins in human tissues, which can be useful for screening samples and validating immunohistochemical results.[35] Using immunohistochemistry images of BC and normal tissues, we investigated differences in protein expression of DEGs between tumors and normal tissues.

**TF-gene Interactions Analysis**

NetworkAnalyst (https://www.networkanalyst.ca/) was applied to find the TF-gene interaction with give hub genes.[36]

**Analysis of Hub Genes with Stemness Score and Immune Cell Infiltration**

A standardized BC dataset was downloaded from the UCSC (https://xenabrowser.net/) database: TCGA and GTEx (BC, N = 1077). Furthermore, we extracted the expression for the ENSG00000253293 (HOXA10) gene in each sample. From previous studies, we were able to obtain six tumor dryness indices based on mRNA expression and methylation characteristics.[37] By using the Spearman's correlation test, we investigated the relationship between the expression of five hub genes and the stemness index of tissue samples containing multiple cancer types. Additionally, we performed log2(x + 0.001) transformations for each expression value and then used the R package ESTIMATE (https://bioinformatics.mdanderson.org/public-software/estimate/) based on gene expression to calculate ImmuneScores (the infiltration of immune cells in tumor tissue), StromalScores (measurement of stromal presence), and EstimateScores (measurement of tumor purity).

**Results**

**Identification of DEGs Between BC and Endometriosis**

Figure 1 illustrates the research flowchart for this study. The datasets GSE5108 and GSE15852 were selected for analysis and identification of the DEGs between BC and endometriosis. Based on the cutoff criteria, we screened 2111 DEGs (1134 upregulated and 977 downregulated) from GSE5108 and 1138
DEGs (506 upregulated and 632 downregulated) from GSE15852 (Fig. 2A, B). Furthermore, the Venn diagram analysis was employed to identify the common DEGs between GSE5108 and GSE15852. In conclusion, we identified 15 overlapping upregulated DEGs (Fig. 2C) and 18 overlapping downregulated DEGs (Fig. 2D). DEGs are listed in Table 1.
Table 1
Detailed information of 33 DEGs.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGHM</td>
<td>immunoglobulin heavy constant mu(IGHM)</td>
</tr>
<tr>
<td>CPM</td>
<td>carboxypeptidase M(CPM)</td>
</tr>
<tr>
<td>DHRS11</td>
<td>dehydrogenase/reductase 11(DHRS11)</td>
</tr>
<tr>
<td>SLC35F2</td>
<td>solute carrier family 35 member F2(SLC35F2)</td>
</tr>
<tr>
<td>FHL2</td>
<td>four and a half LIM domains 2(FHL2)</td>
</tr>
<tr>
<td>HACD2</td>
<td>3-hydroxyacyl-CoA dehydratase 2(HACD2)</td>
</tr>
<tr>
<td>FBXO22</td>
<td>F-box protein 22(FBXO22)</td>
</tr>
<tr>
<td>HOXA10</td>
<td>homeobox A10(HOXA10)</td>
</tr>
<tr>
<td>TRIM9</td>
<td>tripartite motif containing 9(TRIM9)</td>
</tr>
<tr>
<td>EPB41L4B</td>
<td>erythrocyte membrane protein band 4.1 like 4B(EPB41L4B)</td>
</tr>
<tr>
<td>CLMN</td>
<td>calmin(CLMN)</td>
</tr>
<tr>
<td>RGS1</td>
<td>regulator of G protein signaling 1(RGS1)</td>
</tr>
<tr>
<td>MSX1</td>
<td>msh homeobox 1(MSX1)</td>
</tr>
<tr>
<td>SMC04</td>
<td>single-pass membrane protein with coiled-coil domains 4(SMC04)</td>
</tr>
<tr>
<td>PDLIM5</td>
<td>PDZ and LIM domain 5(PDLIM5)</td>
</tr>
<tr>
<td>EMX2</td>
<td>empty spiracles homeobox 2(EMX2)</td>
</tr>
<tr>
<td>FZD5</td>
<td>frizzled class receptor 5(FZD5)</td>
</tr>
<tr>
<td>MME</td>
<td>membrane metalloendopeptidase(MME)</td>
</tr>
<tr>
<td>FZD7</td>
<td>frizzled class receptor 7(FZD7)</td>
</tr>
<tr>
<td>P2RY14</td>
<td>purinergic receptor P2Y14(P2RY14)</td>
</tr>
<tr>
<td>RARRES1</td>
<td>retinoic acid receptor responder 1(RARRES1)</td>
</tr>
<tr>
<td>FN1</td>
<td>fibronectin 1(FN1)</td>
</tr>
<tr>
<td>INHBA</td>
<td>inhibin subunit beta A(INHBA)</td>
</tr>
<tr>
<td>THOC2</td>
<td>THO complex subunit 2(THOC2)</td>
</tr>
<tr>
<td>HOOK2</td>
<td>hook microtubule tethering protein 2(HOOK2)</td>
</tr>
<tr>
<td>DDAH1</td>
<td>dimethylarginine dimethylaminohydrolase 1(DDAH1)</td>
</tr>
</tbody>
</table>
GO and KEGG Pathway Analysis of Overlapping DEGs

For a functional enrichment analysis, 33 DEGs were found to overlap between BC and endometriosis. By entering the up-regulated and down-regulated DEG lists into DAVID, we obtain the corresponding GO and KEGG pathway enrichment analysis. The results of the GO and KEGG enrichment analysis of overlapping DEGs are presented in Table 2,3. As shown in Fig. 3A, 3B, in terms of BP, upregulated DEGs were significantly associated with endodermal cell differentiation, substrate adhesion-dependent cell spreading, positive regulation of cell proliferation, positive regulation of phosphatidylinositol 3-kinase signaling and the downregulated DEGs were principally associated with anterior/posterior pattern specification, negative regulation of cell proliferation, peptide metabolic process, positive regulation of transcription from RNA polymerase II promoter. Based on the analysis of CC results, upregulated DEGs were significantly enriched in blood microparticles, extracellular exosomes, and extracellular space, while downregulated DEGs were significantly enriched in chromatin, transcription factor complex. For MF, the upregulated DEGs were enriched in receptor binding, protein binding involved in cell-cell adhesion, and cell adhesion, whereas the downregulated DEGs were enriched in transcription activator activity and sequence-specific double-stranded DNA binding. In Fig. 3C and 3D, DEGs have significant KEGG pathways related to regulating pluripotency of stem cells and mis-regulating transcription in cancer.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAX8</td>
<td>paired box 8(PAX8)</td>
</tr>
<tr>
<td>PEMT</td>
<td>phosphatidylethanolamine N-methyltransferase(PEMT)</td>
</tr>
<tr>
<td>CD248</td>
<td>CD248 molecule(CD248)</td>
</tr>
<tr>
<td>CD47</td>
<td>CD47 molecule(CD47)</td>
</tr>
<tr>
<td>TCF3</td>
<td>transcription factor 3(TCF3)</td>
</tr>
<tr>
<td>CFB</td>
<td>complement factor B(CFB)</td>
</tr>
<tr>
<td>FGFR1</td>
<td>fibroblast growth factor receptor 1(FGFR1)</td>
</tr>
</tbody>
</table>
Table 2
The analysis of GO and KEGG of upregulated DEGs.

<table>
<thead>
<tr>
<th>Term and Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0009952 (BP)</td>
<td>anterior/posterior pattern specification</td>
</tr>
<tr>
<td>GO:0008285 (BP)</td>
<td>negative regulation of cell proliferation</td>
</tr>
<tr>
<td>GO:0006518 (BP)</td>
<td>peptide metabolic process</td>
</tr>
<tr>
<td>GO:0045944 (BP)</td>
<td>positive regulation of transcription from RNA polymerase II promoter</td>
</tr>
<tr>
<td>GO:0000785 (CC)</td>
<td>chromatin</td>
</tr>
<tr>
<td>GO:0005667 (CC)</td>
<td>transcription factor complex</td>
</tr>
<tr>
<td>GO:0001228 (MF)</td>
<td>transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding</td>
</tr>
<tr>
<td>GO:1990837 (MF)</td>
<td>sequence-specific double-stranded DNA binding</td>
</tr>
<tr>
<td>GO:0000981 (MF)</td>
<td>RNA polymerase II transcription factor activity, sequence-specific DNA binding</td>
</tr>
<tr>
<td>hsa:05202 (KEGG)</td>
<td>Transcriptional mis-regulation in cancer</td>
</tr>
</tbody>
</table>
Table 3
The analysis of GO and KEGG of downregulated DEGs.

<table>
<thead>
<tr>
<th>Term and Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0035987 (BP)</td>
</tr>
<tr>
<td>GO:0034446 (BP)</td>
</tr>
<tr>
<td>GO:0008284 (BP)</td>
</tr>
<tr>
<td>GO:0014068 (BP)</td>
</tr>
<tr>
<td>GO:0072562 (CC)</td>
</tr>
<tr>
<td>GO:0070062 (CC)</td>
</tr>
<tr>
<td>GO:0005615 (CC)</td>
</tr>
<tr>
<td>GO:0005539 (MF)</td>
</tr>
<tr>
<td>GO:0005102 (MF)</td>
</tr>
<tr>
<td>GO:0098632 (MF)</td>
</tr>
<tr>
<td>GO:0098631 (MF)</td>
</tr>
<tr>
<td>hsa:04550 (KEGG)</td>
</tr>
<tr>
<td>hsa:05205 (KEGG)</td>
</tr>
</tbody>
</table>

**PPI Network Construction and Hub Genes Identification**

Using the STRING database, the PPI network was conducted to determine whether there were any interactions among the overlapping DGEs and determine whether there were any connections between BC and endometriosis. Cytoscape software was used to visualize the results. According to Fig. 4A, the PPI network consisted of 23 nodes (11 upregulated DEGs and 12 downregulated DEGs) and 54 edges, with a p value of 0.05. To explore the top 10 important hub genes in the PPI network, we used the MCC, Stress, Betweenness, Radiality, Cloceness, BottleNeck, EPC, Dgree, MNC, and DNNC algorithms in the CytoHubba tools. Finally, five intersection genes were identified, including MSX1, FGFR1, INHBA, HOXA10, and PAX8 (Fig. 4B).

Using GeneMANIA online tools, a co-expression network was constructed to examine the interaction between five hub genes and twenty interacting genes (Fig. 4C). The PPI network of these genes showed a complex network with co-expression of 13.90%, physical interactions of 61.52%, co-localization of 2.79%, predicted of 4.30%, WikiPathways of 9.36%, and pathway of 1.51%. Additionally, we conducted GO and KEGG analyses on five hub genes. Based on the results of the study, hub genes appear to be enriched in regulating transcription from the RNA polymerase II promote (Fig. 4D). Based on KEGG analysis, the hub genes were found to be enriched on the pathways involved in regulating the pluripotency of stem cells and the mis-regulation of these pathways in cancer.
Validation of Expression of Hub Genes in BC

To investigate the association between hub genes and clinical features (pathological stages and age), we analyzed the expression levels of five hub genes using the UALCAN database. As shown in Fig. 5A, lower expression of HOXA10, MSX1 and FGFR1, and higher expression of INHBA in BC compared to normal tissues, with no significant difference in the case of PAX8 (as shown by UALCAN analysis). According to Fig. 5B, HOXA10, PAX8, MSX1, and FGFR1 expression levels in different pathological stages were lower than those in normal sample tissues, while INHBA expression levels were higher. Furthermore, we found that the expression levels of 5 hub genes also differed significantly between age groups (Fig. 5C). In the 21-40-year-old group, HOXA10, PAX8, and MSX1 were significantly down-regulated. The levels of FGFR1 and INHBA expression, however, were lower in younger groups, suggesting a close relationship between age and hub gene expression.

Using the Human Protein Atlas database, we compared the protein expression levels of hub genes (Fig. 6). Notably, in BC and normal breast tissues, some hub genes were not detected. Overall, five hub genes may play a role in regulating breast cancer transcription.

Clinical Prognostic Value Analysis of Hub Genes in BC

In order to determine the clinical prognostic value of hub genes, OS (Overall survival), RFS (Recurrence free survival), and DMFS (Distant Metastasis-Free Survival) of the hub genes were further analyzed using Kaplan-Meier plots. Figure 7A illustrates that HOXA10 mRNA expression was not significantly correlated with clinical outcome. High expression of PAX8 (Fig. 7B) was associated with a significantly higher RFS rate, while high expression of MSX1 (Fig. 7C) was associated with a significantly lower DMFS rate. It is evident from Fig. 7D that low risk of patients with BC had lower expression levels of FGFR1 mRNA, which was associated with poor RFS survival. Furthermore, higher expression levels of INHBA were also associated with unfavorable OS, RFS, and DMFS (Fig. 7E).

TF-gene Interactions Analysis of Hub Genes in BC

The TF gene interaction network (Fig. 8) shows the interaction between three hub genes and TF genes. The results indicate that PAX8, MSX1, and HOXA10 were regulated by six, ten, and forty transcription factors, respectively. An interaction relationship between three critical genes (PAX8, MSX1, and INHBA) was identified as the core component, suggesting a robust mechanism involved in the disease process.

Stemness Score Analysis of Hub Genes in BC

In both hereditary and sporadic breast cancers, stem cell subcomponents are thought to retain the majority of stem cell characteristics, but with malregulated self-renewal pathways that promote tumorigenic differentiation and cell heterogeneity. The cancer stem cell (CSC) is a subpopulation of tumor cells that capable of self-renewal and long-term maintenance of tumors. A genetic and
epigenetic abnormality of dry pathways in cancer cells often results in tumor proliferation and infection of surrounding tissues, resulting in malignant transformation of the tumor. Functional enrichment analysis indicated that hub genes are enriched in pathways of phenotypic regulation of stem cells, therefore we examined the relationship between H and dryness score. Figure 9A illustrates the correlation between hub genes and stem cell scores. Most types of cancer stemness scores were negatively associated with downregulated hub genes, including PAX8, MSX1, and HOXA10, indicating that downregulation of these genes may activate the dry pathway, resulting in cell overgrowth and malignant characteristics.

**Discussion**

Known as an estrogen-dependent benign disease, endometriosis can cause pelvic pain, infertility, and even cancers.\[40\] However, several risk factors, including family aggregation, genetic mutations or polymorphisms, and environmental toxins, have received considerable attention from scholars.\[41\] Though endometriosis is a benign disease, it exhibits biological and behavioral characteristics similar to malignant tumors, including local and distant metastasis, attachment, injury, and invasion.\[42, 43\] There are currently several mechanisms by which endometriosis contributes to the development of ovarian cancer, including oxidative stress, inflammatory processes mediated by cytokines and mediators associated with the endometriosis environment, and the increased levels of estrogen observed in the disease.\[44\] A number of studies have shown specific mutations in ARID1A, PIK3CA, PTEN, and other genes to be associated with endometriosis and cancers.\[45–50\] Despite this, the mechanism behind this correlation remains unclear.

BC is the most common malignancy among women worldwide.\[2\] The most effective means of improving the survival rate of BC are early diagnosis and timely treatment. A number of factors may contribute to the development of BC, including genetic factors, reproductive factors, hormones, and so on.\[51\] Epidemiological studies have reported conflicting results regarding the relationship between endometriosis and breast cancer.\[52–54\] It is possible that BC and endometriosis share similar pathogenic pathways. Various growth factors, signaling molecules, and transcription factors are induced by hormones under the influence of mammary glands to regulate growth and development.\[55\] As a result,
exploring molecular mechanisms linking endometriosis to breast cancer, as well as early identification and intervention, is undoubtedly of great clinical importance. In spite of this, no study has been conducted to examine the shared genetic mechanisms and biomarkers of these two diseases.

Bioinformatic analyses of the endometriosis and BC databases were performed in this study. There were 33 DEGs that overlapped, of which 15 were upregulated and 18 were downregulated. In addition, enrichment of GO and KEGG pathways was conducted to investigate the underlying mechanisms of those DEGs. In our study, we found that these pathways are mainly involved in transcriptional regulation, cell proliferation, and PI3K signaling pathways. It is possible for tumors to be initiated either by transformed differentiated cells or by transformed tissue-resident stem cells.[56, 57] Currently, breast cancer is viewed as a stem cell disease, which means that there are cancer cells with stem cell characteristics and tumor-initiation potential, and these cells are responsible for initiation and metastasis of tumors.[58] In order for cancer to develop, stationary stem cells must be driven to become non-renewable cancer cells with unlimited proliferation potential.[59] Recent years have seen an increase in attention given to the stem cell origin theory of endometriosis. A growing number of studies have demonstrated that endometrial stem cells play a role in the development of endometriosis, despite the absence of direct evidence to the contrary.[60, 61] A common pathogenesis of endometriosis and breast cancer may be stem cell regulation, according to our results.

Furthermore, the hub genes were screened by the PPI network. Among the DEGs, a total of five DEGs (HOXA10, FGFR1, MSX1, PAX8, and INHBA) were identified as potential hub genes. As a member of the homeobox gene family, HOXA10 plays an important role in cell proliferation, apoptosis, metabolism, and migration.[62] In addition to its role in regulating embryonic morphogenesis and differentiation, HOXA10 is aberrantly expressed in most types of cancers.[63–65] In spite of the fact that high levels of HOXA10 expression have been shown to promote the proliferation and invasion of cancer cells, it appears to have the opposite effect on breast cancer.[66–68] In our study, HOXA10 was shown to be downregulated in BC tissues, particularly in advanced BC tissues. Additionally, HOXA10 has been implicated in the development of endometriosis. Studies have shown that ectopic endometrium contains less HOXA10 than normal endometrium, which is in agreement with our results.[69] According to a recent study, deficiency of HOXA10 can cause endometrial hyperplasia and progress to endometrial cancer.[70] The results of this study suggest that HOXA10 may play an important role in the progression of BC and endometriosis.

A member of the transcription factor gene pairing box (PAX) family, PAX8 can be used to distinguish gynecological malignancies from non-gynecological malignancies.[71, 72] Among the functions of PAX8 in tumor development are epigenetic remodeling, stimulation of proliferation, inhibition of apoptosis, and regulation of angiogenesis.[73] Furthermore, PAX8 is an epithelial marker highly sensitive to extragenital endometriosis.[74] PAX8 has been used as a biomarker to distinguish metastatic ovarian cancer from breast cancer.[75] Currently, there is limited understanding of the mechanism by which PAX8 contributes to
breast cancer. The present study found that PAX8 was downregulated in BC tissues, particularly in early BC, which comprised the previous study.[71]

As a member of the muscle segment homeobox gene family, MSH homeobox 1 (MSX1) is a transcriptional repressor that plays a key role in embryonic development.[76, 77] There has been an association between abnormal methylation of MSX1 promoter DNA and lung cancer, stomach cancer, ovarian cancer, and breast cancer.[78] Dysregulation of MSX1 expression contributes to various cellular processes in cancer, such as proliferation, invasion, metastasis, tumor dryness, angiogenesis, and so on. Dysregulation of MSX-1 downregulation is important for the regulation of endometrial receptivity.[82, 83] There have been no studies involving MSX1 and endometriosis to date. Despite this, MSX1 was significantly downregulated in endometriosis compared to normal endometrium in our study. Accordingly, MSX1 may play an important role in the progression of BC and endometriosis, although further biological investigation is necessary.

Many physiological processes are mediated by fibroblast growth factor receptors (FGFRs), including embryonic and postnatal development, metabolic homeostasis, and cancer progression.[84] A previous study demonstrated that FGFR1 was commonly overexpressed in ectopic endometrium of endometriosis compared with its eutopic counterpart or normal endometrium, suggesting that it may play a role in endometriosis.[85] In addition, FGFR1 plays a crucial role in BC.[86] The expression of FGFR1 was lower in BC tissues than in normal tissues. FGFR1 amplification, however, was significantly associated with the induction of cancer cell stemness, metastasis, and drug resistance in BC.[87–89] Therefore, we hypothesized that FGFR1 plays a significant role in BC and endometriosis.

Inhibin β-A (INHBA) belongs to the superfamily of transforming growth factors (TGFs). INHBA acts either as a cancer suppressor or a cancer promoter according to the cellular environment.[90] In gastric cancer and Wilms tumor, INHBA significantly inhibited tumor growth and angiogenesis.[91] Overexpression of INHBA, however, is associated with poor prognoses for gastric, esophageal, and lung cancer.[92–94] Currently, it is widely believed that INHBA plays an important role in the renewal and pluripotency of embryonic stem cells.[95, 96] INHBA induces epithelial-mesenchymal transformation (EMT) and accelerates the movement of BC cells by activating the TGF-β regulatory gene, which plays a key role in promoting the proliferation and invasion of BC cells.[90] A high level of INHBA expression has also been associated with breast cancer aggressiveness.[97] Similar to the findings of these studies, INHBA concentrations are markedly higher in BC tissues than in normal tissues, especially in advanced BC, suggesting that INHBA may play an important role in the proliferation of BC. A number of studies have demonstrated that INHBA is highly expressed in the endometrium and can act as a local regulator of differentiation, growth, and invasion of human cytotrophoblast cells.[98, 99]

Women of childbearing age are susceptible to endometriosis, as we all know. Based on age-stratified statistics of breast cancer patients, we found that the expression of 3 down-regulated hub genes is higher in the 21–40 age group than in other age groups, whereas the expression of 2 up-regulated hub genes is...
lower. Based on these results, it appears that BC and endometriosis may share a common initiation mechanism in women, especially young women.

Tumor microenvironment (TME) is a complex cellular environment composed of adaptive or innate immune cells, mesenchymal cells, endothelial cells, inflammatory mediators, and extracellular matrix molecules.\cite{100} The TME plays an important role in the progression and prognosis of malignant tumors.\cite{101} TME consists primarily of infiltrating stromal cells and immune cells. This study calculated three immune-related scores for five hub genes, which showed that ESTIMATEScore, ImmuneScore, and StromalScore were positively correlated with HOXA10 expression, while FGFR1 was negatively correlated. According to these findings, hub genes may play a role in the regulation of disease progression within the immune microenvironment.

There were no studies that explored the hub genes and common molecular mechanisms associated with BC and endometriosis. The potential correlation between BC and endometriosis led us to explore and identify for the first time common DEGs and hub genes, which may assist in clarifying the common mechanisms. Therefore, it may be useful for diagnosing and treating BC and endometriosis at an early stage.

There are, however, some limitations in our research. The study was conducted retrospectively, focusing only on the microarray expression cohort, and further validation in laboratory experiments is required. Additionally, we validated the DEGs in our study only in patients with BC or endometriosis due to the lack of a dataset including patients with both BC and endometriosis. As a third point, the PPI and most of the molecular mechanisms of genes in this study were exclusively based on predictions from public databases, which will need to be further examined in future studies.

**Conclusion**

In conclusion, by integrating multiple microarray gene expression profiles, 5 hub genes (HOXA10, MSX1, PAX8, FGFR1, and INHBA) were identified that may be related to the occurrence and development of BC and endometriosis. Endometriosis complicated with BC is revealed in this study to provide new insights into the molecular mechanisms of this disease.

**Declarations**

**Acknowledgments**

The authors thank the staff of the Tianjin Medical University Cancer Institute & Hospital for providing helpful insights and suggestions.

**Author Contributions**
All authors participated in preparing the manuscript and approved the submitted version. Linyue Hai were responsible for experimental design, experimental analysis and thesis writing. Xuchen Cao and Chunhua Xiao were responsible for the guidance and review of the thesis.

**Funding**

This study was funded by Tianjin Key Medical Discipline (Specialty) Construction Project TJYXZDXK-009A and Wu Jieping Medical Foundation (320.6750.2021-10-2).

**Availability of data and materials**

Publicly available datasets were analyzed in this study. This data can be found here: All the raw data used in this study are derived from the public GEO data portal (https://www.ncbi.nlm.nih.gov/geo/; Accession numbers: GSE15852 and GSE5108). Further inquiries can be directed to the corresponding author.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

**References**


Figures

HOXA10, PAX8, MSX1, FGFR1, INHBA

clinical prognostic value of hub genes

stemness and immune score of hub genes
The flow chart of this research.

Figure 2

Identification of DEGs between BC and endometriosis. The volcano map of GSE5108 (A) and GSE15852 (B). The upregulated genes are marked in light red and downregulated genes are marked in blue green. The venn diagram of 15 upregulated DEGs (C) and 18 downregulated DEGs (D). DEGs were determined using GEO2R software. FDR < 0.05 and FC > 1.5 were considered as the cutoff values.
Figure 3

Functional enrichment analysis of DEGs. (A) Significantly enriched GO and KEGG pathways of the upregulated DEGs. (B) Significantly enriched GO and KEGG pathways of the downregulated DEGs. (C) Identification of significantly enriched GO and KEGG clusters by using Metascape enrichment analysis. The GO enriched terms are colored by p-value. (D) An interactive network of the top enrichment pathways.
Distinct circle nodes represented GO or KEGG terms. The size of nodes was proportional to the number of consensus genes related to the terms. Only terms with a similarity score of >0.3 were connected by edges.

Figure 4

Identification of hub genes and construction of PPI network. (A) PPI network of DEGS constructed by STRING and Cytoscape. The blue nodes represented upregulated DEGs and the red nodes represented
downregulated DEGs. The edges indicated interconnections between different DEGs. (B) The 5 hub genes based on the MCC, Stress, Betweenness, Radiality, Cloceness, BottleNeck, EPC, Dgree, MNC, and DNNC algorithms in CytoHubba tools were identified. (C) The hub genes and their co-expression genes were analyzed by GeneMANIA. The different colors of the network edge indicate the distinct bioinformatics methods. (D) The GO and KEGG analysis of 5 hub genes were presented as bubble maps.

Figure 5
(A) The expression level of 5 hub genes between normal and BC tissues. (B) The expression level of 5 hub genes in different stages of BC. (C) The expression of hub genes in different age groups.

Figure 6

HOXA10, PAX8, MSX1, and FGFR1 proteins expression in normal and BC tissues from the HPA database.
Figure 7

Clinical prognostic value analysis of hub genes in BC. Kaplan–Meier analysis of OS, RFS, and DMFS in TCGA liver cancer dataset based on HOXA10 (A), PAX8 (B), MSX1 (C), FGFR1 (D), and INHBA (E) expression.
Figure 8

TF-gene regulatory network.
Figure 9

Stemness score and immune cell infiltration analysis of hub genes in BC. (A) Correlation between the expression of hub genes and stemness score based on DNAss, DMPss, ENHss, RNAss, EREG-METHss, and EREG.EXPss. (B) The scatter plots of correlation between hub genes and stromal score, immune score, ESTIMATE score in BC.